

# **Petite chronique de l'évaluation : évoluer pour mieux évaluer ?**

**Esas - 8 janvier 2015**

Pascal Detroz, IFRES-ULg

# Durant cette conférence nous ne parlerons pas ...

- Des différents modèles de l'évaluation
- Des différentes fonctions de l'évaluation
- Des aspects sociologiques, philosophiques et politiques de l'évaluation
- Des aspects psychologiques et éthiques de l'évaluation
- Des évaluations externes
- De l'évaluation des Hautes Ecoles

# **I. Une définition de l'évaluation**

# Conceptualisation de l'évaluation

- Evaluer c'est recueillir de l'information sur laquelle sera posée un jugement afin de prendre des décisions (Stufflebeam, 1981)

## **2. L'humain est-il un bon évaluateur ?**

# Biais liés à nos perceptions

**Count how many times  
the players wearing  
white pass the ball**

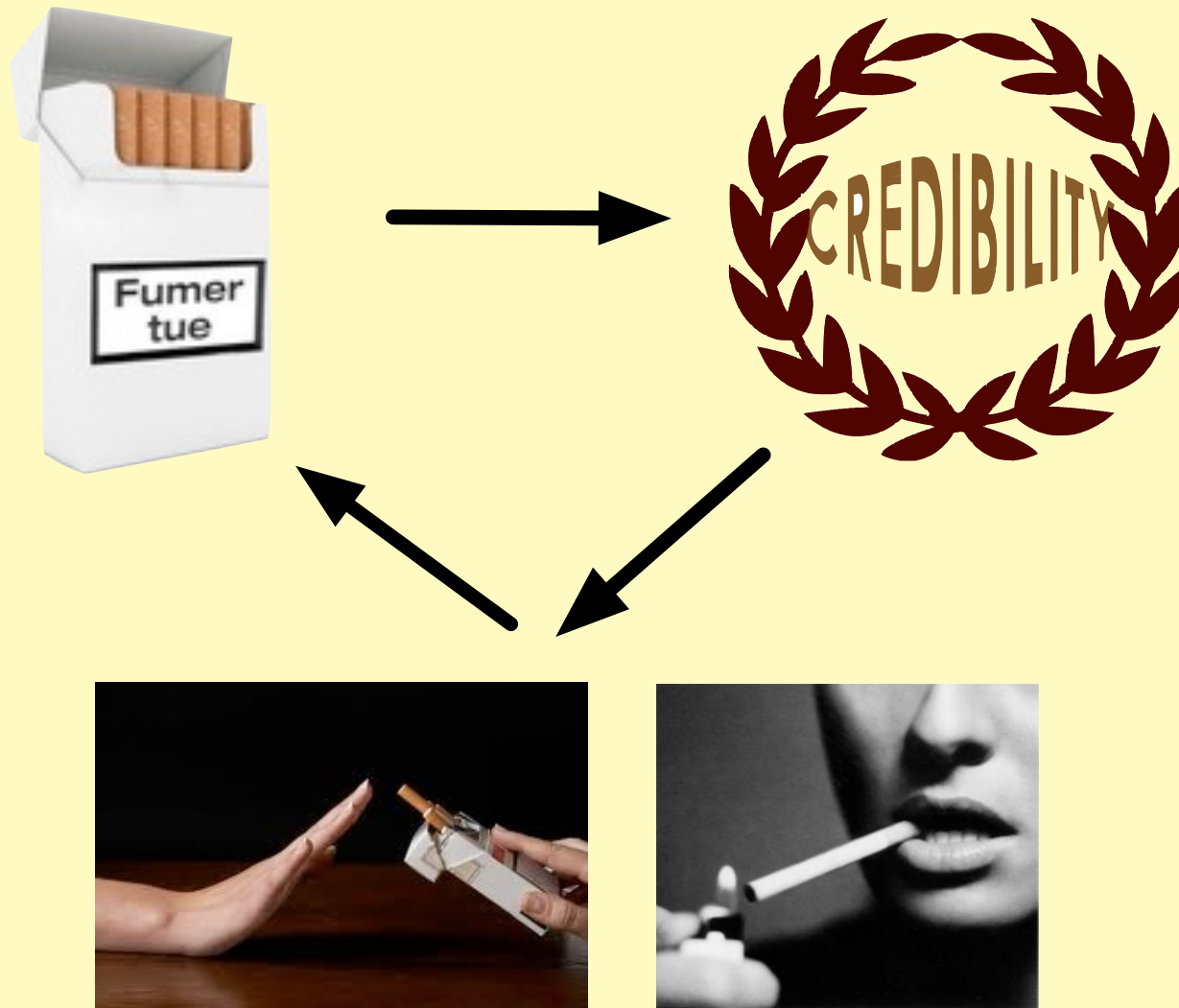
# Biais liés aux interprétations

## ➤ Mis en évidence par la psychologie sociale

Exemple : l'égotisme implicite. Nous accordons plus d'importance et nous ressentons plus d'attraction pour les personnes, les lieux et les activités qui contiennent les lettres de notre nom et les chiffres de notre date de naissance (Koole et al., 2001 ; Pelhal et al., 2002 et 2011 ; Jones et al.s 2004).

Exemple : The Totalitarian Ego (Greenwald, 1980). Le besoin d'un concept de soi stable est si intense qu'il agit comme un dictateur qui contrôle l'information et réécrit l'histoire à son avantage personnel.

# Biais liés à l'action

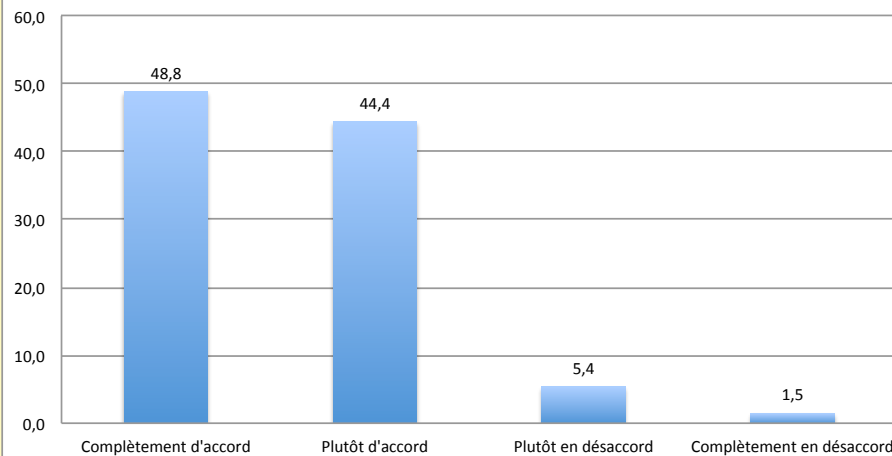




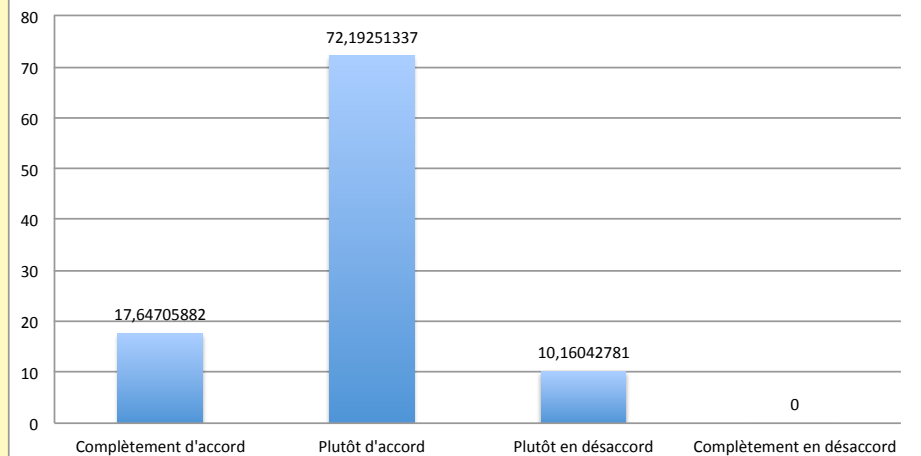
# Est-on conscient de nos lacunes

- Nous sommes des enseignants !  
Nous savons évaluer !

**Evaluer les étudiants nécessite des connaissances et des savoir-faire liés au domaine de l'évaluation**



**Mes connaissances en évaluation me permettent de bien évaluer les étudiants**



# Des biais dans NOS évaluations ?

- ❖ Un même jeu de copies corrigées plusieurs fois par un même correcteur sans qu'il ne s'en rende compte
- ❖ Un même jeu de copies corrigées par plusieurs correcteurs différents
- ❖ Une même copie est placée dans un ensemble de copies dans des positions différentes
- ❖ Une même copie est placée dans un ensemble de copies dont les valeurs sont plus ou moins dispersées largement
- ❖ Une même copie est corrigée par plusieurs groupes de correcteurs auxquels on fournit des informations complémentaires différentes sur l'élève

# Des biais dans NOS évaluations ?

- Nous sommes des enseignants universitaires !  
Nous savons évaluer !

Agazzi (1967). Pour six domaines, si correcteurs notent une série d'examen (sur 20).

Le seuil de réussite est à 10. voici les résultats :

	6 notes insuffisantes	Avis discordant	6 notes suffisantes
Version latine	0,4	0,5	0,1
Composition Française	0,21	0,7	0
Anglais	0,37	0,47	0,16
Mathématique	0,44	0,36	0,2
Philosophie	0,09	0,81	0,1
Physique	0,37	0,5	0,13

# Des biais dans NOS évaluations ?

## Résultats de recherche :

- La qualité formelle des copies
- L'orthographe et la grammaire
- La longueur du texte
- Texte manuscrit VS texte dactylographié
- L'effet de contraste ou de contexte
- La fidélité intra-correcteur

# Des biais dans NOS évaluations ?

## Résultats de recherche :

- L'expérience du correcteur
- L'état transitoire du correcteur
- La fidélité inter-correcteurs
- Le genre de l'évalué
- L'ethnie de l'étudiant
- L'effet de halo sur l'origine sociale des étudiants

# Des biais dans NOS évaluations ?

Résultats de recherche :

- L'effet de stéréotypie ou d'inertie
- La distribution forcée (loi de posthumus)

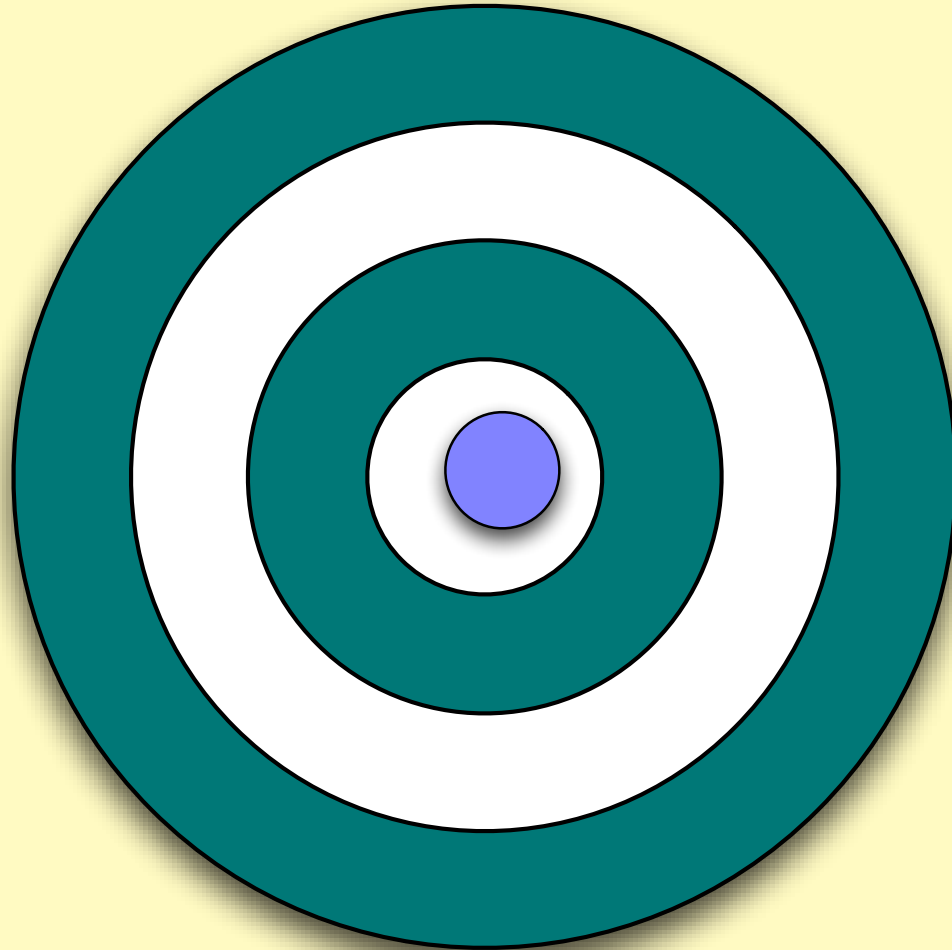
# Conclusions provisoires

- Evaluer les élèves est quelque chose de complexe qui doit être pris au sérieux et nécessite une compétence scientifique (docimologique) forte. Contrairement à notre intuition, nous sommes des évaluateurs très subjectifs, surtout quand nous sommes engagés d'un point de vue émotionnel, ce qui est le cas quand nous évaluons les étudiants qui sont nos élèves.
- Nous avons besoin d'outils et de méthodes !

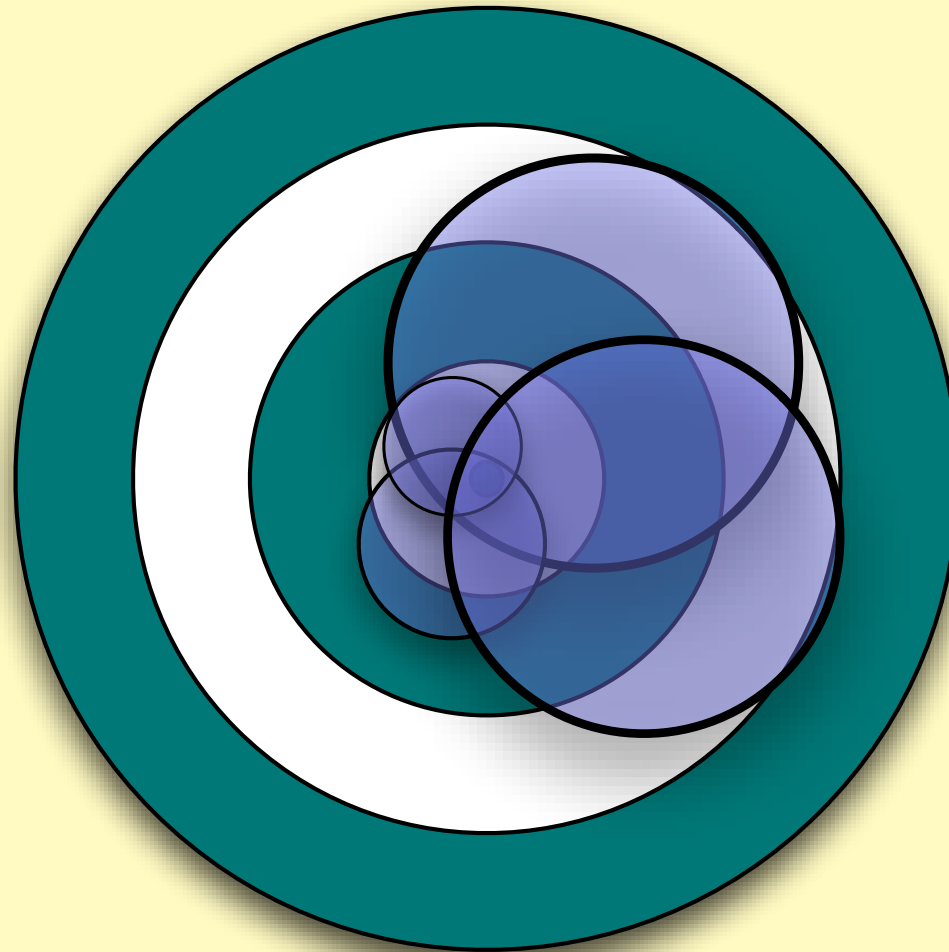
### **3. Le problème de la mesure**



# La mesure physique



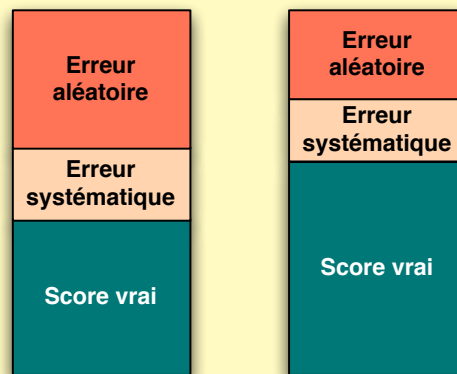
# La mesure en Sciences Humaines



Une nécessaire triangulation

# Une autre représentation

- Le score observé (par exemple à lors de l'évaluation des enseignements par les étudiants) est toujours la somme d'un score vrai et d'une erreur de mesure
- Cette erreur peut être systématique ou aléatoire
- Le but de l'évaluateur est de diminuer la part d'erreur de mesure et d'augmenter la part du score vrai



# Le concept de validité (AERA, APA, NCME, 1999)

- Ce concept est défini comme l'adéquation et la pertinence des inférences faites à partir des résultats d'un test
- Cinq sources de preuve doivent être prises en compte dans les études de validité (le contenu, le processus de réponse, la structure interne du test, la relation avec d'autres variables mesurées, les conséquences du test)
- La validité n'est pas inhérente au test décontextualisé

## 4. Le seuil de césure

## **4.1 Le seuil de césure : enjeu théorique**

# Le problème

La fonction des examens est d'estimer la compétence des individus dans une matière donnée.

Or, la compétence (au sens large) est une disposition personnelle permanente à réussir certaines tâches ou certaines activités. Cette disposition personnelle et permanente est par nature hors d'atteinte de l'investigation directe : elle est invisible et intangible, au contraire de concepts physiques comme le poids et la taille.

# Le problème

La seule connaissance que l'on peut en avoir est indirecte, à travers l'observation des performances de l'étudiant, c'est-à-dire de ce qui est réellement accompli dans un contexte particulier : l'examen.

C'est donc à travers un artefact, le dispositif d'évaluation, que nous allons recueillir une/des mesure(s) de la performance et que nous allons inférer que celle(s)-ci reflète(nt) la compétence.



# Le problème

Dans un monde idéal, dans lequel la performance reflète exactement la compétence, nous n'aurions pas besoin de seuil de césure. Les étudiants compétents répondraient correctement à l'ensemble des items.

Mais nous ne sommes pas dans un monde idéal.

1. L'artefact - le dispositif d'évaluation - n'est souvent pas à même de donner une mesure parfaitement valide de la performance (conception, passation, correction).

2. Les inférences seraient possibles si les tâches/items du test (permettant de mesurer la performance) étaient représentatifs de l'univers des tâches/items attestant de la compétence évaluée.

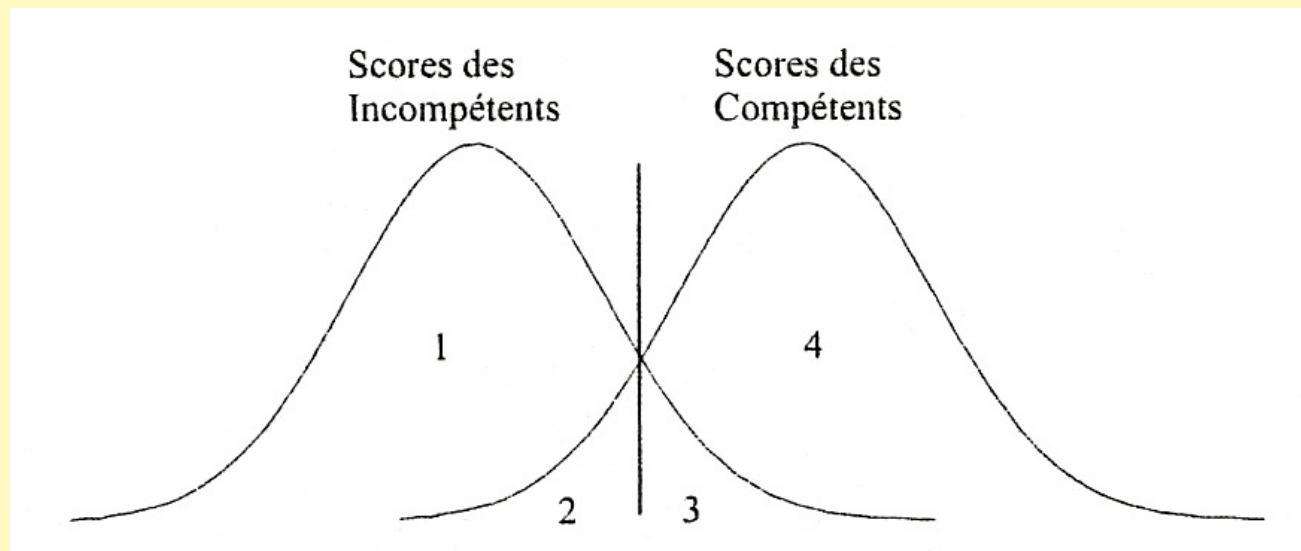
# Le problème

En conséquence, il est acceptable qu'un étudiant, pourtant suffisamment compétent, ne réussisse pas l'ensemble des tâches/items d'une épreuve évaluative. On peut donc attester de la compétence de quelqu'un n'ayant pas satisfait à l'ensemble des items.

Considérant une épreuve destinée à estimer la compétence des individus qui y sont soumis - via des mesures de performances - et formée d'un certain nombre d'items qui peuvent individuellement être réussis ou ratés et fournissant chacun un point de score, à quel niveau faut-il fixer le score en-dessous duquel la compétence n'est pas attestée?

# Les enjeux de cette question

Si on avait une mesure de la compétence qui nous permettait de déterminer la population des « compétents » et celle des « incompétents », le croisement de cette mesure avec la performance observée représentée par le score à l'examen nous donnerait quatre profils : les compétents-performants, les compétents non-performants, les incompétents-performants, les incompétents-non-performants.



# Que faire si on n'a pas de mesure préalable de la compétence ?

**Théoriquement** : appliquer la méthode des jugements de questions.

Cette méthode nécessite un jugement collectif des questions par une série d'experts ou de pairs de l'enseignant. La notion collective est importante car cette méthode s'appuie sur l'ambition de limiter l'arbitraire de l'enseignant isolé, doublement souverain des questions qu'il pose et du score de coupure qu'il impose.

La question à laquelle doivent répondre les experts est, pour chaque tâche/question : est-il possible qu'un étudiant compétent échoue à cette question ? Et si oui, dans quelle proportion ? A partir de là, il devient possible de considérer le seuil minimal nécessaire pour attester de l'acquisition de la compétence.

# Que faire si on n'a pas de mesure préalable de la compétence, ni de panel d'experts sous la main ?

Pratiquement :

On fixe le seuil de manière arbitraire (d'abord 12, puis maintenant 10) et on dit aux enseignants : « Tirez votre plan ! »

Ce qui n'empêche pas le débat (Cf. CEB en primaire)

## 4.2. Le seuil de césure en pratique

# Les enseignants qui « tirent leur plan »

L'hypothèse socio-arithmétique de Hutchmacher (1993)

Rot et Butas (1959) décrivent l'expérience de Gjorgjevski. Dans celle-ci, cinq professeurs doivent noter cent copies. Il extrait ensuite 15 copies ayant obtenu le même score et les redistribue à quatre autres enseignants. La dispersion des notes entre les deux phases de correction est identique. C'est l'effet de tendance centrale.

Pinot de Moreira (2001) : certains correcteurs ajustent leur sévérité au fur et à mesure des corrections passant de sévères à cléments et vice-versa.

Farrell et Gilbert (1960) démontrent que l'attribution des notes varie en fonction du nombre de copies corrigées. Selon leur étude, les notes extrêmes (très négatives ou très positives) sont attribuées en fin de correction.

Conséquence : la distribution forcée (loi de Posthumus)



## **4.3. La courbe en cloche, l'amie infidèle**



# Chouette, une distribution normale

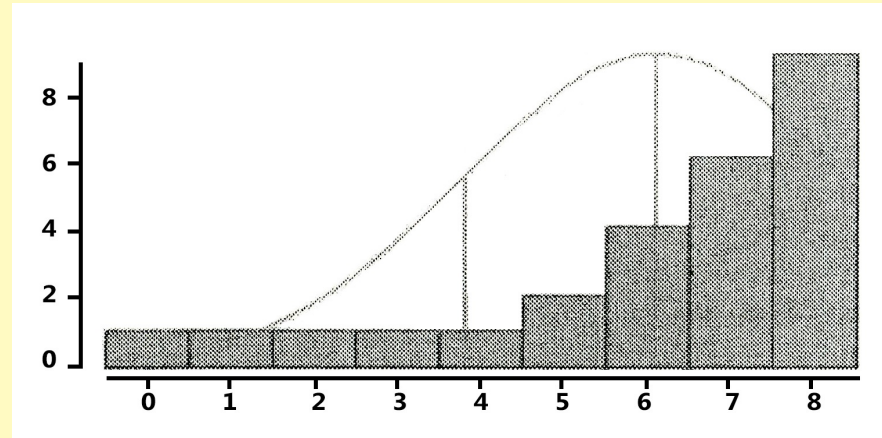
La loi normale, qui rend compte de beaucoup de phénomènes naturels, est considérée pour cette raison comme l'expression même de l'ordre naturel des choses. Idéologiquement elle est rassurante. Sociologiquement, elle évite les ennuis.

Techniquement, elle décrit des phénomènes dont l'amplitude dépend de l'addition d'un grand nombre de phénomènes indépendants et de faible ampleur. Autrement dit : l'accumulation d'un tas de petites choses fortuites. Ou encore l'expression du hasard, conçu comme le constat désolé de l'ignorance où nous sommes des causes détaillées de ce qui se passe autour de nous.

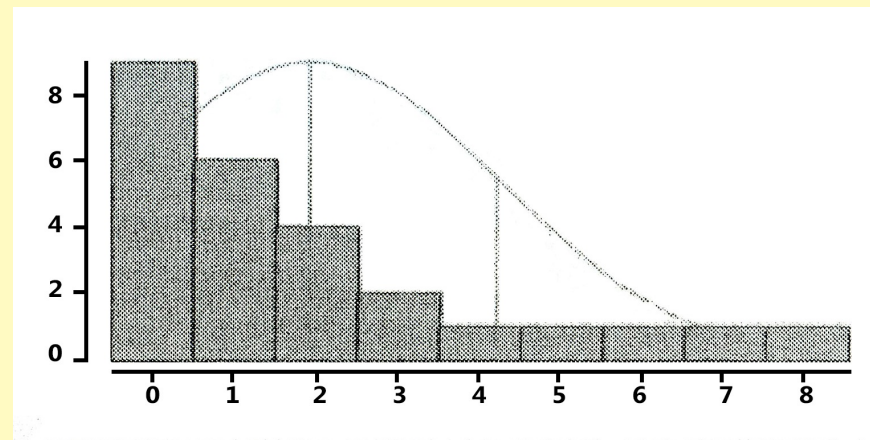
Or, l'enseignant, idéalement, lorsqu'il mesure la compétence, n'est pas dans un état d'incertitude complet par rapport à ce qu'il cherche à mesurer. Son rôle est même de bien pouvoir en éclairer les mécanismes d'apprentissage : de faire passer ses étudiants d'un état réputé incompetent vers un état de compétence.

## D'autres types de courbes ?

En début de formation, il serait vraisemblable d'obtenir une courbe en L



En fin d'apprentissage, il serait tout aussi vraisemblable d'obtenir une courbe en J



Si l'on obtient une courbe normale, quelle a été l'efficacité de notre enseignement ?

## En résumé

*Dans le cas où l'on pense posséder une certaine maîtrise et compréhension du fonctionnement des phénomènes mesurés, et de leur évolution, comme ce devrait être le cas dans les situations d'évaluation en éducation, ce qu'on mesure n'est pas dû au hasard, et l'apparition d'une distribution normale des résultats est tout à fait anormale au sens usuel (Alain Dubus, 2006).*

Antibi (2003) parle même de constante macabre : la croyance en l'universalité de la loi normale en situation d'évaluation scolaire constitue une distorsion du jugement amenant à concevoir et corriger des épreuves évaluatives qui mènent à la réplique de telle courbe.

La loi de Posthumus, clairement dénoncée par les études PISA, nous amène à dire que quels que soient les efforts pédagogiques réalisés, il n'est pas certain qu'ils diminueront le taux d'échec. Les étudiants seront sans doute plus compétents, mais nos dispositifs d'évaluation seront incapables de notifier ces différences.

## **4.4. Et nous, concrètement, on fait quoi ?**

# Le score à l'examen

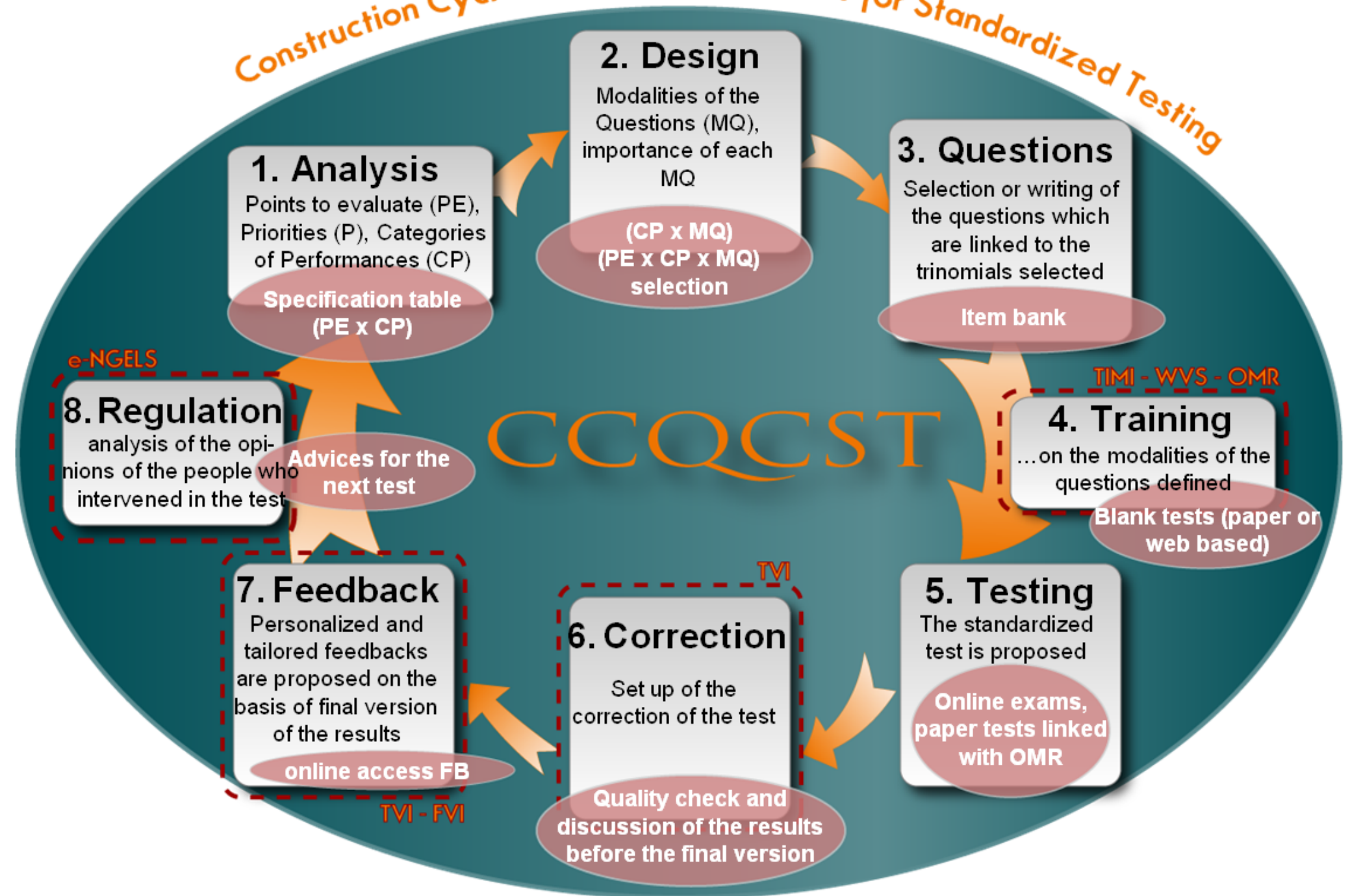
Il est dépendant de quatre éléments :

- La compétence de l'étudiant
- Les facteurs transitoires qui peuvent influencer la performance de l'étudiant
- La difficulté du test
- La sévérité de la correction et les barèmes utilisés

Au vu de ce que nous venons de présenter, il n'y a pas grand chose à craindre pour les oraux et les écrits. L'enseignant, habitué à « tirer son plan » et soucieux de laisser passer ceux qui ont les compétences minimales pour « aller plus loin », va réajuster plus ou moins inconsciemment la difficulté du test ou la sévérité de la correction. Ou, plus probablement, opérer un glissement d'échelle.

## **5. Vers un cycle qualité en évaluation**

# Construction Cycle and Quality Controls for Standardized Testing



# Etape I : Analyse

Content Outline	Recalling information taught or read	Applying knowledge in situations very similar to those taught	Applying knowledge in a new or novel context
<b>I. Basic Parts of Cell</b>	<b>1</b>	<b>1</b>	<b>1</b>
A.Nucleus			
B Cytoplasm			
C. Cell Membrane			
<b>II. Plant vs Animal Cells</b>	<b>2</b>	<b>0</b>	<b>0</b>
A. Similarities			
B. Differences			
1. Cell wall vs membrane			
2. food manufacture			
<b>III. Cell Membrane</b>	<b>2</b>	<b>2</b>	<b>0</b>
A. Living nature of			
B. Diffusion			
C. Substances diffused by cells			
<b>IV. Division of Cells</b>	<b>4</b>	<b>1</b>	<b>0</b>
A. Phases in Division			
B. Chromosomes and DNA			
C. Plant vs animal cell division			



# Etape I : Analyse

Taxonomie des apprentissages cognitifs de Bloom revue par Pohl (1999)					
Connaitre	Comprendre	Appliquer	Analyser	Evaluer	Créer
répéter, mémoriser, définir, connaître, énumérer, rappeler, écrire	discuter, reformuler, traduire, décrire, reconnaître	utiliser, employer, développer, resoudre, traduire, interpréter, appliquer	différencier, identifier, trouver, analyser, reconnaître, explorer, comparer	Juger, estimer, évaluer, composer	Composer, planifier, proposer, désigner, formuler, produire, réaliser

# Exemple : Table de spécification

Examen écrit : Employé d'administration D6

			Catégorie de performance				
			Connaître	Comprendre	Appliquer	Analyser	PRIORITE
Chapitre I : Déontologie	<b>I</b>	<b>Les devoirs</b>					
	<b>II</b>	<b>L'évaluation des agents</b>					
	<b>III</b>	<b>Les incompatibilités</b>					
	<b>IV</b>	<b>Le régime disciplinaire</b>					
	IV.1.	Sanctions					
	IV.2.	Procédure					
	IV.3.	Autorité disciplinaire					
Chapitre II : Principes généraux du droit public belge	<b>I</b>	<b>Le droit et ses subdivisions</b>					
	<b>II</b>	<b>La Constitution</b>					
	II.1.	Définition					
	II.2.	Révision de la Constitution					
	II.3.	Révisions intervenues depuis 1831					
	II.4.	Régime des libertés constitutionnelles					
	<b>III</b>	<b>Les caractéristiques de l'Etat belge</b>					
	III.1.	La séparation des pouvoirs					
	III.2.	La Belgique est une monarchie constitutionnelle et héréditaire					
	III.3.	La Belgique est une démocratie représentative et parlementaire					
	III.4.	La Belgique est un Etat de droit					
	III.5.	La Belgique est un Etat fédéral					
	<b>IV</b>	<b>La hiérarchie des normes</b>					
	<b>V</b>	<b>Le territoire et ses subdivisions</b>					
	V.1.	Définition et limites					
	V.2.	Les subdivisions					
		- Les régions linguistiques					
		- Les communautés					
		- Les régions					
		- Les provinces					
		- Les communes					
	<b>VI</b>	<b>Notions de droit administratif</b>					
		- Le principe de la continuité du service public					
		- Le principe du changement					
		- Le principe de l'égalité des usagers					

## Etape 2 : Design

### Choix des modalités et options de questionnement

- Choisir les modalités en s'appuyant sur les éléments de la table de spécification
- Multiplier les méthodes d'évaluation
- Evaluer à l'oral les performances les plus complexes
- L'oral prend du temps, favorise l'erreur aléatoire, est souvent faible en terme de validité de contenu, rend difficile la traçabilité
- L'oral permet l'évaluation de compétences langagières, peut être diagnostique et formatif, favorise la validité de processus de réponse

# Etape 2 : Le design de l'évaluation

- Il existe énormément de modalités de questions
  - Vrai-Faux.
  - QCM.
  - Phrase à compléter.
  - QROC.
  - QROM.
  - Résolution de problèmes.
  - QROL.
  - Démonstration.
  - ...
- Chacune avec de nombreuses options
  - Evaluation orale ou écrite.
  - A livre ouvert ou fermé.
  - Individuelle ou collective.
  - Avec temps imparti ou non.
  - ...

## Etape 2 : Le design de l'évaluation

Chaque type de question a ses avantages et ses inconvénients. Par exemple, les QCM permettent de poser de nombreuses questions en un temps limité, mais ne mesurent pas les performances les plus complexes.

Le choix entre les modalités et options de questionnement doit tenir compte des :

- objectifs d'apprentissages visés (et uniquement de ceux-ci);
- importances relatives des objectifs d'apprentissages visés (cfr table de spécification);
- décisions qui doivent être nourries par le processus d'évaluation;
- principes de faisabilité et de pragmatisme.

# Etape 3 : Item

## Etape de création du dispositif d'évaluation

- Poser des questions en liens avec la table de spécification
- Soigner la rédaction des questions et des attendus
- Standardiser les méthodes d'évaluation
- Vers un dispositif de type Ecos
- Réfléchir à la traçabilité du dispositif

## Etape 3 : Questions

- C'est l'étape où les questions sont créées en référence à la table de spécification.
- L'enjeu principal est : cet item évalue-t-il avec validité au moins un des objectifs d'apprentissages les plus importants référencés dans la table de spécification ?
- L'enjeu est également de construire une banque de questions pérenne.
- Différents critères d'analyse formelle s'appliquent aux différentes modalités de questionnement et à leurs options.

# Etape 3: Une checklist pour vérifier la qualité des QROC

- L'item est-il écrit sous forme de question ?
- L'item est-il rédigé de manière suffisamment claire telle que la réponse correcte apparait clairement comme faisant une courte phrase, un mot unique ou un chiffre ?
- Le blanc (ou la réponse attendue) apparait-il à la fin de la phrase ?
- L'item est-il paraphrasé (en comparaison avec un contenu copié-collé du cours) ?
- Si l'item est sous forme de phrase à compléter, le mot attendu est-il bien un concept central plutôt qu'une référence anecdotique ?
- Le blanc (ou l'espace laissé pour la réponse) fait-il systématiquement la même taille ?



# Etape 3: Une checklist pour vérifier la qualité des QCM

- L'amorce pose-t-elle une question directe ou cerne-t-elle bien un problème spécifique ?
- L'item porte-t-il bien sur un corpus théorique plus large qu'un simple directement issu des notes de cours ?
- Sauf si c'est strictement nécessaire, le vocabulaire et la structure de la phrase sont-ils simples à comprendre
- Chacun des distracteurs est-il plausible ?
- Les distracteurs et la réponse correcte sont-ils formellement équivalents ?
- Si possible, les distracteurs représentent-ils une erreur ou une incompréhension classique de la part des étudiants ?
- La réponse correcte est-elle indépendante des distracteurs, de l'amorce, ou d'autres items du questionnaire ?
- Existe-t-il une et une seule bonne réponse ?

# Etape 3: Une checklist pour vérifier la qualité des productions longues

- La production longue nécessite-elle que l'étudiant applique ses compétences à une situation qu'il n'a pas rencontrée explicitement dans le cours ?
- Quand on la compare aux autres items du test, la production longue contribue-t-elle à couvrir des objectifs pédagogiques spécifiés dans la table de spécification.
- La question est-elle posée de manière spécifique ? L'étudiant sait-il indubitablement la direction que l'on attend qu'il prenne ?
- La tâche demandée est-elle compatible avec le degré de maturité/maturation des étudiants ?
- L'étudiant doit-il utiliser dans sa réponse plus que des catégories de performance de bas niveau ?
- Les consignes et/ou la question précisent-ils clairement
  - l'amplitude et la longueur de la production attendue ?
  - l'objectif que vise la question et qu'il doivent atteindre ?
  - le temps qu'il est raisonnable de consacrer à la réponse à cette question ?
  - les critères d'évaluation auxquels cette question sera confrontée ?

# Etape 4 : Information

Buts : communiquer aux étudiants...

- les informations sur le test (quand, comment, avec quelles modalités d'évaluation, sous quelles conditions, ...)
- ce que comprend le test et sur quoi il se concentre particulièrement
- des outils leur permettant de s'entraîner à fournir les performances attendues
- des moyens de gérer leur anxiété

# Etape 4: Quelques conseils aux étudiants

- Stratégie de gestion du temps
  - Commencer à travailler de manière productive tout de suite.
  - Passer les questions pour lesquelles ils ne savent quoi répondre (pour y revenir plus tard).
  - La cas échéant, utiliser les temps disponibles à la fin du test pour relire les questions.
- Mettre en place une stratégie évitant les erreurs grossières.
  - Prêter attention aux consignes et à la question afin de déterminer clairement la nature de la tâche qui est attendue.
  - Ne pas hésiter à demander des clarifications pendant l'examen (si c'est autorisé - pas sur le contenu).
  - Vérifier sur le fond et la forme, chacune des réponses données.
- Pour les QCM, adapter sa stratégie au barème utilisé et utiliser, en cas de doute, un raisonnement déductif.

# Etape 4 : La meilleure manière d'entraîner les étudiants : le test à blanc

- Organiser un test à blanc similaire à celui qui sera utilisé lors de l'examen. Si possible, il se structurera autour des :
  - mêmes objectifs d'apprentissages;
  - mêmes types de documents (consigne, feuille de réponse,...);
  - mêmes conditions de temps;
  - mêmes modalités de questionnement;
  - mêmes degrés de difficulté;
  - mêmes procédures anti-fraudes;
  - mêmes conditions exogènes.
  - ...

# Etape 5 : Testing

Veiller à

- tenir compte des rapports de force
- protéger les étudiants de nous-mêmes
- garder une attitude professionnelle
- standardiser au mieux les procédures d'évaluation

## Etape 6 : correction

Travailler à l'aide de grille d'évaluation pour sérier les informations, les mesurer, les pondérer, mais aussi pour assurer la traçabilité du jugement et de la décision.

# Etape 6: Correction

- C'est une étape très technique qui comprend 3 phases essentielles :
  - [1] évaluer la production de l'étudiant;
  - [2] effectuer des contrôles qualité sur la correction et/ou les items;
  - [3] effectuer d'éventuelles modifications subséquentes;
- Pour chaque modalité de questionnement, il existe différentes modalités de correction et différents types de contrôle qualité.



# Etape 6 : Correction des QCM

- L'évaluation de la production de l'étudiant est automatique.
- Des décisions doivent être prise préalablement sur le barème.
- Les contrôles qualité portent sur les items via le “r.bis”.
  - En tendance, les étudiants les meilleurs au test ont-ils choisi la réponse correcte alors que ce n'est pas le cas des étudiants les plus faibles au test ? Observe-t-on l'effet inverse pour les distracteurs ?
  - Sinon que faire ? Vérifier l'alerte !
  - Si elle est confirmée, il existe la possibilité :
    - de valoriser un des autres distracteurs;
    - de modifier la réponse correcte;
    - d'éliminer l'une des questions.
- Possibilité de contrôle qualité sur le test.

# Etape 6 : Correction des QROC

- L'évaluation de la performance de l'étudiant est plus ou moins automatique.
  - On peut anticiper la plupart des réponses considérées comme correctes qui sont attendues.
  - Mais parfois, cela n'empêche pas le jugement subjectif (par exemple, que fait on avec les réponses correctes, mais mal orthographiées ?)
- Le barème doit être défini préalablement.
- Les contrôles qualité portent sur les items via le "r.bis".
  - En tendance, les étudiants les meilleurs au test ont-ils choisi la réponse correcte alors que ce n'est pas le cas des étudiants les plus faibles au test ? Observe-t-on l'effet inverse pour les distracteurs ?
  - Sinon que faire ? Vérifier l'alerte !
  - Si elle est confirmée, il existe la possibilité :
    - de valoriser un des autres distracteurs;
    - de modifier la réponse correcte;
    - d'éliminer l'une des questions.
- Possibilité de contrôle qualité sur le test.

# Etape 6 : correction des productions longues : les échelles d'évaluation

- Les échelles d'évaluation métriques : traduisent un jugement portant sur la qualité en une échelle métrique

CRITERIA OF DRAWING EVALUATION

<u>8</u>	TITLE BLOCK	<i>take more time</i>
<u>9</u>	LINE TECHNIQUE	<i>&amp; care on your title block</i>
<u>10</u>	CENTERING AND SPACING	<i>lettering.</i>
<u>—</u>	ARCS, CIRCLES, TANGENTS	
<u>—</u>	SPACING OF DIMENSIONS	
<u>—</u>	PLACEMENT OF DIMENSIONS	
<u>7</u>	FRACTIONS, FIGURES, LETTERING	
<u>—</u>	ARROWHEADS	
<u>9+</u>	NEATNESS, OVERALL APPEARANCE	
<u>10</u>	SOLUTION	

TOTAL 53 AVERAGE \_\_\_\_\_ GRADE B+

*HORIZON*

*nice guidelines!*

*nice choice of subject*

# Etape 6 : correction des productions longues : les échelles d'évaluation

- Les échelles graphiques ordinales : un continuum délimité et échelonné par des descripteurs de qualité

Teacher or observer:

---

**Directions:** Place a check mark any place along the line to show judgment of the student's performance on that item. If you have not had sufficient opportunity to observe this student, circle N/O.

---

**ACHIEVEMENT OF GROUP GOALS**

1. Does the student attend the group meetings?

Never	Seldom	Occasionally	Frequently	Always	N/O
-------	--------	--------------	------------	--------	-----

2. When attending is the student prepared?

Never	Seldom	Occasionally	Frequently	Always	N/O
-------	--------	--------------	------------	--------	-----

3. Does the student work actively toward achieving the group's goals?

Never	Seldom	Occasionally	Frequently	Always	N/O
-------	--------	--------------	------------	--------	-----

4. Does the student work outside of the group meetings on the group project?

Never	Seldom	Occasionally	Frequently	Always	N/O
-------	--------	--------------	------------	--------	-----

# Etape 6 : correction des productions longues : les échelles d'évaluation

- Les échelles descriptives : Similaire aux précédentes, sauf que l'on remplace les formulations ambiguës (souvent, parfois, ...) par des descriptions du comportement attendu pour valoriser un niveau donnée.

Student's name: Topic:	Date:		
1. Did the student clearly state the thesis or main point?			
0	1	2	3
Did not state or imply the main point or thesis	Implied the main point or thesis but did not state it clearly	Stated the main idea or thesis clearly but only matter-of-factly	Stated the main idea or thesis clearly, enthusiastically, and interestingly for the audience
2. Did the student define the key terms when necessary to do so?			
0	1	2	NA
No attempt to define key terms, even when it was necessary to do so	Attempts to define the key terms, but was not effective in doing so	Clearly and effectively defines the necessary key terms	The presentation was such that defining key terms was unnecessary
3. Did the student use sound reasoning to support the main point or thesis?			
0	1	2	3
Offered no supporting reasons for the thesis or position taken	Supporting reasons given but they are off-target or they do not lend direct support for the thesis	Gives relevant supporting reasons, but could have given better or more diverse reasons	Gives excellent supporting reasons, good diversity, directly applicable

# Etape 6 : Améliorer l'utilisation des échelles descriptives

- Organiser les diverses rubriques de l'échelle descriptive de manière logique et en référence à la table de spécification.
- Pour chaque niveau de l'échelle descriptive, décrire minutieusement les comportements attendus.
- Fournir des exemples issus des productions d'étudiants pour illustrer les comportements attendus.
- Travailler en collaboration pour produire les échelles descriptives.
- Faire valider les échelles descriptives par des experts contenu.
- Entraîner et superviser toutes les personnes qui vont utiliser les échelles descriptives.
- Faire corriger une même copie par plusieurs correcteurs.
- Monitorer les corrections et vérifier la fidélité intra et inter-correcteurs.

# Etape 6 : Checklist pour évaluer le processus d'évaluation

- Les rubriques de l'échelle mettent-elles l'emphasis sur les contenus les plus importants de la table de spécification ?
- Les étudiants comprennent-ils les rubriques ?
- Les rubriques et leur correction offrent-elles à l'étudiant une information et un feedback susceptibles de les amener à améliorer leur performance en lien avec les items évalués ?
- Les échelons des différentes échelles sont-ils clairement décrits en terme de performance réellement observable ?
- Les échelons des différentes échelles permettent-ils de distinguer les différences de niveaux entre les étudiants ?

## Etape 7 : Feedback

Donner un feedback diagnostique et complet aux étudiants sur leurs performances



# Etape 7: feedback

- Rappelons que l'évaluation est en lien avec un processus de décision
- Assessment provides students, among others, with information to make decisions about learning (Nitko, 2001).
- Les bons feedbacks sont descriptifs, spécifiques et contiennent de l'information susceptible d'améliorer les performances de l'étudiant.
- Il existe trois type de feedbacks
  - des feedbacks normatifs : on compare la performance d'un étudiant à celle des autres
  - des feedbacks autoréférencés : on compare la production actuelle d'un étudiant à sa production passée.
  - des feedbacks critériés : on compare la production d'un étudiant à des standards établis en spécifiant les objectifs d'apprentissage atteints de ceux qui ne le sont pas.

# Etape 7: feedback et table de spécification

On peut donner un feedback sur un item

Content Outline	Recalling information taught or read	Applying knowledge in situations very similar to those taught	Applying knowledge in a new or novel context
<b>I. Basic Parts of Cell</b>	<b>1</b>	<b>1</b>	<b>1</b>
A.Nucleus		1 item (label parts) of cell drawing	
B Cytoplasm	1 item (short-answer)		
C. Cell Membrane			1 item (label part of cell photo)
<b>II. Plant vs Animal Cells</b>	<b>2</b>	<b>0</b>	<b>0</b>
A. Similarities			
B. Differences			
1. Cell wall vs membrane	1 item (short-answer)		
2. food manufacture	1 item (short-answer)		
<b>III. Cell Membrane</b>	<b>2</b>	<b>2</b>	<b>0</b>
A. Living nature of	1 item (short-answer)	1 item (multiple choice)	
B. Diffusion		1 item (multiple choice)	
C. Substances diffused by cells	1 item (short-answer)		
<b>IV. Division of Cells</b>	<b>4</b>	<b>1</b>	<b>0</b>
A. Phases in Division	2 item (short-answer)		

# Etape 7: feedback et table de spécification

On peut donner un feedback relatif à un objectif d'apprentissage spécifique

Content Outline	Recalling information taught or read	Applying knowledge in situations very similar to those taught	Applying knowledge in a new or novel context
<b>I. Basic Parts of Cell</b>	<b>1</b>	<b>1</b>	<b>1</b>
A. Nucleus	1 item (short-answer)	1 item (label parts) of cell drawing	1 item (label part of cell photo)
B. Cytoplasm			
C. Cell Membrane			
<b>II. Plant vs Animal Cells</b>	<b>2</b>	<b>0</b>	<b>0</b>
A. Similarities	1 item (short-answer)		
B. Differences			
1. Cell wall vs membrane			
2. food manufacture	1 item (short-answer)		
<b>III. Cell Membrane</b>	<b>2</b>	<b>2</b>	<b>0</b>
A. Living nature of	1 item (short-answer)	1 item (multiple choice)	
B. Diffusion	1 item (short-answer)	1 item (multiple choice)	
C. Substances diffused by cells			
<b>IV. Division of Cells</b>	<b>4</b>	<b>1</b>	<b>0</b>
A. Phases in Division	2 item (short-answer)		

# Etape 7: feedback et table de spécification

On peut donner un feedback relatif à un objectif d'apprentissage général

Content Outline	Recalling information taught or read	Applying knowledge in situations very similar to those taught	Applying knowledge in a new or novel context
<b>I. Basic Parts of Cell</b>	<b>1</b>	<b>1</b>	<b>1</b>
A. Nucleus	1 item (short-answer)	1 item (label parts) of cell drawing	1 item (label part of cell photo)
B. Cytoplasm			
C. Cell Membrane			
<b>II. Plant vs Animal Cells</b>	<b>2</b>	<b>0</b>	<b>0</b>
A. Similarities	1 item (short-answer)		
B. Differences			
1. Cell wall vs membrane			
2. food manufacture	1 item (short-answer)		
<b>III. Cell Membrane</b>	<b>2</b>	<b>2</b>	<b>0</b>
A. Living nature of	1 item (short-answer)	1 item (multiple choice)	
B. Diffusion	1 item (short-answer)	1 item (multiple choice)	
C. Substances diffused by cells			
<b>IV. Division of Cells</b>	<b>4</b>	<b>1</b>	<b>0</b>
A. Phases in Division	2 item (short-answer)		

# Etape 7: feedback et table de spécification

On peut donner un feedback relatif à un élément de matière spécifique

Content Outline	Recalling information taught or read	Applying knowledge in situations very similar to those taught	Applying knowledge in a new or novel context
<b>I. Basic Parts of Cell</b>	<b>1</b>	<b>1</b>	<b>1</b>
A. Nucleus	1 item (short-answer)	1 item (label parts) of cell drawing	1 item (label part of cell photo)
B. Cytoplasm			
C. Cell Membrane			
<b>II. Plant vs Animal Cells</b>	<b>2</b>	<b>0</b>	<b>0</b>
A. Similarities	1 item (short-answer)		
B. Differences			
1. Cell wall vs membrane			
2. food manufacture	1 item (short-answer)		
<b>III. Cell Membrane</b>	<b>2</b>	<b>2</b>	<b>0</b>
A. Living nature of	1 item (short-answer)	1 item (multiple choice)	
B. Diffusion	1 item (short-answer)	1 item (multiple choice)	
C. Substances diffused by cells			
<b>IV. Division of Cells</b>	<b>4</b>	<b>1</b>	<b>0</b>
A. Phases in Division	2 item (short-answer)		

# Etape 7: feedback et table de spécification

On peut donner un feedback relatif à un élément de matière spécifique

Content Outline	Recalling information taught or read	Applying knowledge in situations very similar to those taught	Applying knowledge in a new or novel context
<b>I. Basic Parts of Cell</b>	<b>1</b>	<b>1</b>	<b>1</b>
A. Nucleus	1 item (short-answer)	1 item (label parts) of cell drawing	1 item (label part of cell photo)
B. Cytoplasm			
C. Cell Membrane			
<b>II. Plant vs Animal Cells</b>	<b>2</b>	<b>0</b>	<b>0</b>
A. Similarities	1 item (short-answer)		
B. Differences			
1. Cell wall vs membrane			
2. food manufacture	1 item (short-answer)		
<b>III. Cell Membrane</b>	<b>2</b>	<b>2</b>	<b>0</b>
A. Living nature of	1 item (short-answer)	1 item (multiple choice)	
B. Diffusion	1 item (short-answer)	1 item (multiple choice)	
C. Substances diffused by cells			
<b>IV. Division of Cells</b>	<b>4</b>	<b>1</b>	<b>0</b>
A. Phases in Division	2 item (short-answer)		

# Etape 7: feedback et table de spécification

On peut donner un feedback relatif à un élément de matière général

Content Outline	Recalling information taught or read	Applying knowledge in situations very similar to those taught	Applying knowledge in a new or novel context
<b>I. Basic Parts of Cell</b>	<b>1</b>	<b>1</b>	<b>1</b>
A.Nucleus	1 item (short-answer)	1 item (label parts) of cell drawing	1 item (label part of cell photo)
B Cytoplasm			
C. Cell Membrane			
<b>II. Plant vs Animal Cells</b>	<b>2</b>	<b>0</b>	<b>0</b>
A. Similarities	1 item (short-answer)		
B. Differences			
1. Cell wall vs membrane			
2. food manufacture	1 item (short-answer)		
<b>III. Cell Membrane</b>	<b>2</b>	<b>2</b>	<b>0</b>
A. Living nature of	1 item (short-answer)	1 item (multiple choice)	
B. Diffusion	1 item (short-answer)	1 item (multiple choice)	
C. Substances diffused by cells			
<b>IV. Division of Cells</b>	<b>4</b>	<b>1</b>	<b>0</b>
A. Phases in Division	2 item (short-answer)		

# Etape 7: feedback et table de spécification

On peut donner un feedback relatif à un niveau taxonomique

Content Outline	Recalling information taught or read	Applying knowledge in situations very similar to those taught	Applying knowledge in a new or novel context
<b>I. Basic Parts of Cell</b>	<b>1</b>	<b>1</b>	<b>1</b>
A. Nucleus	1 item (short-answer)	1 item (label parts) of cell drawing	1 item (label part of cell photo)
B. Cytoplasm			
C. Cell Membrane			
<b>II. Plant vs Animal Cells</b>	<b>2</b>	<b>0</b>	<b>0</b>
A. Similarities	1 item (short-answer)		
B. Differences			
1. Cell wall vs membrane			
2. food manufacture	1 item (short-answer)	<b>2</b>	<b>0</b>
<b>III. Cell Membrane</b>	<b>2</b>		
A. Living nature of	1 item (short-answer)		
B. Diffusion	1 item (short-answer)	1 item (multiple choice)	
C. Substances diffused by cells		1 item (multiple choice)	
<b>IV. Division of Cells</b>	<b>4</b>	<b>1</b>	<b>0</b>
A. Phases in Division	2 item (short-answer)		



# Etape 8 : Rétroaction

Vers une spirale qualité

**Merci de votre attention !**